October 2005

Relative compositionality of multi-word expressions: a study of verb-noun (V-N) collocations

Sriram Venkatapathy
International Institute of Information Technology

Aravind K. Joshi
University of Pennsylvania, joshi@cis.upenn.edu

Postprint version. Published in *Lecture Notes in Computer Science*, Volume 3651, Natural Language Processing (IJCNLP 2005), pages 553-564.
Publisher URL: http://dx.doi.org/10.1007/11562214_49

Relative compositionality of multi-word expressions: a study of verb-noun (V-N) collocations

Abstract
Recognition of Multi-word Expressions (MWEs) and their relative compositionality are crucial to Natural Language Processing. Various statistical techniques have been proposed to recognize MWEs. In this paper, we integrate all the existing statistical features and investigate a range of classifiers for their suitability for recognizing the non-compositional Verb-Noun (V-N) collocations. In the task of ranking the V-N collocations based on their relative compositionality, we show that the correlation between the ranks computed by the classifier and human ranking is significantly better than the correlation between ranking of individual features and human ranking. We also show that the properties 'Distributed frequency of object' (as defined in [27]) and 'Nearest Mutual Information' (as adapted from [18]) contribute greatly to the recognition of the non-compositional MWEs of the V-N type and to the ranking of the V-N collocations based on their relative compositionality.

Comments
Postprint version. Published in *Lecture Notes in Computer Science*, Volume 3651, Natural Language Processing (IJCNLP 2005), pages 553-564.
Publisher URL: http://dx.doi.org/10.1007/11562214_49

Relative compositionality of multi-word expressions: a study of verb-noun (V-N) collocations

Sriram Venkatapathy[1] and Aravind K. Joshi[2]

[1] Language Technologies Research Center,
International Institute of Information Technology - Hyderabad,
Hyderabad, India.
sriram@research.iiit.ac.in.
[2] Department of Computer and Information Science and
Institute of Research in Cognitive Science,
University of Pennsylvania, Philadelphia, PA, USA.
joshi@linc.cis.upenn.edu

Abstract. Recognition of Multi-word Expressions (MWEs) and their relative compositionality are crucial to Natural Language Processing. Various statistical techniques have been proposed to recognize MWEs. In this paper, we integrate all the existing statistical features and investigate a range of classifiers for their suitability for recognizing the non-compositional Verb-Noun (V-N) collocations. In the task of ranking the V-N collocations based on their relative compositionality, we show that the correlation between the ranks computed by the classifier and human ranking is significantly better than the correlation between ranking of individual features and human ranking. We also show that the properties 'Distributed frequency of object' (as defined in [27]) and 'Nearest Mutual Information' (as adapted from [18]) contribute greatly to the recognition of the non-compositional MWEs of the V-N type and to the ranking of the V-N collocations based on their relative compositionality.

1 Introduction

The main goals of the work presented in this paper are **(1)** To investigate a range of classifiers for their suitability in recognizing the non-compositional V-N collocations, and **(2)** To examine the relative compositionality of collocations of V-N type. Measuring the relative compositionality of V-N collocations is extremely helpful in applications such as machine translation where the collocations that are highly non-compositional can be handled in a special way.

Multi-word expressions (MWEs) are those whose structure and meaning cannot be derived from their component words, as they occur independently. Examples include conjunctions like 'as well as' (meaning 'including'), idioms like

[1] Part of the work was done at Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA 19104, USA, when he was visiting IRCS as a visiting Scholar, February to December, 2004.

'kick the bucket' (meaning 'die'), phrasal verbs like 'find out' (meaning 'search') and compounds like 'village community'. A typical natural language system assumes each word to be a lexical unit, but this assumption does not hold in case of MWEs [6] [12]. They have idiosyncratic interpretations which cross word boundaries and hence are a 'pain in the neck' [23]. They account for a large portion of the language used in day-to-day interactions [25] and so, handling them becomes an important task.

A large number of MWEs have a standard syntactic structure but are non-compositional semantically. An example of such a subset is the class of non-compositional verb-noun collocations (V-N collocations). The class of V-N collocations which are non-compositional is important because they are used very frequently. These include verbal idioms [22], support-verb constructions [1] [2] etc. The expression 'take place' is a MWE whereas 'take a gift' is not a MWE.

It is well known that one cannot really make a binary distinction between compositional and non-compositional MWEs. They do not fall cleanly into mutually exclusive classes, but populate the continuum between the two extremes [4]. So, we rate the MWEs (V-N collocations in this paper) on a scale from 1 to 6 where 6 denotes a completely compositional expression, while 1 denotes a completely opaque expression. But, to address the problem of identification, we still need to do an approximate binary distinction. We call the expressions with a rating of 4 to 6 compositional and the expressions with rating of 1 to 3 as non-compositional. (See Section 4 for further details).

Various statistical measures have been suggested for identification of MWEs and ranking expressions based on their compositionality. Some of these are Frequency, Mutual Information [9], Log-Likelihood [10] and Pearson's χ^2 [8]. Integrating all the statistical measures should provide better evidence for recognizing MWEs and ranking the expressions. We use various Machine Learning Techniques (classifiers) to integrate these statistical features and classify the V-N collocations as MWEs or Non-MWEs. We also use a classifier to rank the V-N collocations according to their compositionality. We then compare these ranks with the ranks provided by the human judge. A similar comparison between the ranks according to Latent-Semantic Analysis (LSA) based features and the ranks of human judges has been done by McCarthy, Keller and Caroll [19] for verb-particle constructions. (See Section 3 for more details). Some preliminary work on recognition of V-N collocations was presented in [28].

In the task of classification, we show that the technique of weighted features in distance-weighted nearest-neighbour algorithm performs slightly better than other machine learning techniques. We also find that the 'distributed frequency of object (as defined by [27])' and 'nearest mutual information (as adapted from [18])' are important indicators of the non-compositionality of MWEs. In the task of ranking, we show that the ranks assigned by the classifier correlated much better with the human judgement than the ranks assigned by individual statistical measures.

This paper is organised in the following sections **(2)** Basic Architecture, **(3)** Related work, **(4)** Data used for the experiments, **(5)** Agreement between

the Judges, **(6)** Features, **(7)** Experiments - Classification, **(8)** Experiments - Ranking and **(9)** Conclusion.

2 Basic Architecture

Recognition of MWEs can be regarded as a classification task where every V-N collocation can be classified either as a MWE or as a Non-MWE. Every V-N collocation is represented as a vector of features which are composed largely of various statistical measures. The values of these features for the V-N collocations are extracted from the British National Corpus. For example, the V-N collocation 'raise an eyebrow' can be represented as
[Frequency = 271, Mutual Information = 8.43, Log-Likelihood = 1456.29, etc.].

Now, to recognise the MWEs, the classifier has to do a binary classification of this vector. So, ideally, the classifier should take the above information and classify 'raise an eyebrow' as an MWE. The classifier can also be used to rank these vectors according to their relative compositionality.

3 Related Work

Church and Hanks (1989) proposed a measure of association called Mutual Information [9]. Mutual Information (MI) is the logarithm of the ratio between the probability of the two words occurring together and the product of the probability of each word occurring individually. The higher the MI, the more likely are the words to be associated with each other. The usefulness of the statistical approach suggested by Church and Hanks [9] is evaluated for the extraction of V-N collocations from German text Corpora [7]. Several other measures like Log-Likelihood [10], Pearson's χ^2 [8], Z-Score [8] , Cubic Association Ratio (MI3), Log-Log [17], etc., have been proposed. These measures try to quantify the association of the two words but do not talk about quantifying the non-compositionality of MWEs. Dekang Lin proposes a way to automatically identify the non-compositionality of MWEs [18]. He suggests that a possible way to separate compositional phrases from non-compositional ones is to check the existence and mutual-information values of phrases obtained by replacing one of the words with a similar word. According to Lin, a phrase is probably non-compositional if such substitutions are not found in the collocations database or their mutual information values are significantly different from that of the phrase. Another way of determining the non-compositionality of V-N collocations is by using 'distributed frequency of object'(DFO) in V-N collocations [27]. The basic idea in there is that "if an object appears only with one verb (or few verbs) in a large corpus we expect that it has an idiomatic nature" [27].

Schone and Jurafsky [24] applied Latent-Semantic Analysis (LSA) to the analysis of MWEs in the task of MWE discovery, by way of rescoring MWEs extracted from the corpus. An interesting way of quantifying the relative compositionality of a MWE is proposed by Baldwin, Bannard, Tanaka and Widdows [3]. They use latent semantic analysis (LSA) to determine the similarity between

an MWE and its constituent words, and claim that higher similarity indicates great decomposability. In terms of compositionality, an expression is likely to be relatively more compositional if it is decomposable. They evaluate their model on English NN compounds and verb-particles, and showed that the model correlated moderately well with the Wordnet based decomposibility theory [3].

Evert and Krenn [11] compare some of the existing statistical features for the recognition of MWEs of adjective-noun and preposition-noun-verb types. Galiano, Valdivia, Santiago and Lopez [14] use five statistical measures to classify generic MWEs using the LVQ (Learning Vector Quantization) algorithm. In contrast, we do a more detailed and focussed study of V-N collocations and the ability of various classifiers in recognizing MWEs. We also compare the roles of various features in this task.

McCarthy, Keller and Caroll [19] judge compositionality according to the degree of overlap in the set of most similar words to the verb-particle and head verb. They showed that the correlation between their measures and the human ranking was better than the correlation between the statistical features and the human ranking. We have done similar experiments in this paper where we compare the correlation value of the ranks provided by the classifier with the ranks of the individual features for the V-N collocations. We show that the ranks given by the classifier which integrates all the features provides a significantly better correlation than the individual features.

4 Data used for the experiments

The data used for the experiments is British National Corpus of 81 million words. The corpus is parsed using Bikel's parser [5] and the Verb-Object Collocations are extracted. There are 4,775,697 V-N of which 1.2 million were unique. All the V-N collocations above the frequency of 100 (n=4405) are taken to conduct the experiments so that the evaluation of the system is feasible. These 4405 V-N collocations were searched in Wordnet, American Heritage Dictionary and SAID dictionary (LDC,2003). Around 400 were found in at least one of the dictionaries. Another 400 were extracted from the rest so that the evaluation set has roughly equal number of compositional and non-compositional expressions. These 800 expressions were annotated with a rating from 1 to 6 by using guidelines independently developed by the authors. 1 denotes the expressions which are totally non-compositional while 6 denotes the expressions which are totally compositional. The brief explanation of the various rating are **(1)** No word in the expression has any relation to the actual meaning of the expression. Example : "**leave a mark**". **(2)** Can be replaced by a single verb. Example : "**take a look**". **(3)** Although meanings of both words are involved, at least one of the words is not used in the usual sense. Example : "**break news**". **(4)** Relatively more compositional than (3). Example : "**prove a point**". **(5)** Relatively less compositional than (6). Example : "**feel safe**". **(6)** Completely compositional. Example : "**drink coffee**". For the experiments on classification (Section 7), we call the expressions with ratings of 4 to 6 as compositional and the expressions

with rating of 1 to 3 as non-compositional. For the experiments on ranking the expressions based on their relative compositionality, we use all the 6 ratings to represent the relative compositionality of these expressions.

5 Agreement between the Judges

The data was annotated by two fluent speakers of English. For 765 collocations out of 800, both the annotators gave a rating. For the rest, atleast one of the annotators marked the collocations as "don't know". Table 1 illustrates the details of the annotations provided by the two judges.

Ratings	6	5	4	3	2	1	Compositional	Non-Compositional
							(4 to 6)	(1 to 3)
Annotator1	141	122	127	119	161	95	390	375
Annotator2	303	88	79	101	118	76	470	195

Table 1. Details of the annotations of the two annotators

From the table we see that annotator1 distributed the rating more uniformly among all the collocations while annotator2 observed that a significant proportion of the collocations were completely compositional. To measure the agreement between the two annotators, we used the Kendall's TAU (τ). τ is the correlation between the rankings[1] of collocations given by the two annotators. W ranges between 0 (little agreement) and 1 (full agreement). W is calculated as below,

$$\tau = \frac{\sum_{i<j} sgn(x_i - x_j) sgn(y_i - y_j)}{\sqrt{(T_0 - T_1)(T_0 - T_2)}}$$

where $T_0 = n(n-1)/2$, $T_1 = \sum t_i(t_i - 1)/2$, $T_2 = \sum u_i(u_i - 1)/2$ and where, n is the number of collocations, t_i is the number of tied x values of i^{th} group of tied x values and u_i is the number of tied y values of i^{th} group of tied y values.

We obtained a τ score of **0.61** which is highly significant. This shows that the annotators were in a good agreement with each other in deciding the rating to be given to the collocations. We also compare the ranking of the two annotators using Spearman's Rank-Correlation coefficient (r_s) (more details in section 8). We obtained a r_s score of **0.71** indicating a good agreement between the annotators. A couple of examples where the annotators differed are **(1)** "perform a task" was rated 3 by annotator1 while it was rated 6 by annotator2 and **(2)** "pay tribute" was rated 1 by annotator1 while it was rated 4 by annotator2.

The 765 samples annotated by both the annotators were then divided into a training set and a testing set in several possible ways to cross-validate the results of classification and ranking.

[1] computed from the ratings

Feature	Top-3		Feature	Top-3
Frequency	take place have effect have time		Mutual Information [9]	shrug shoulder bridge gap plead guilty
Cubic Association Measure (Oakes, 1998)	take place shake head play role		Log-Log [17]	shake head commit suicide fall asleep
Log-Likelihood [10]	take place shake head play role		Pearson's χ^2 [8]	shake head commit suicide fall asleep
T-Score [9]	take place have effect shake head		Z-Score [26]	shake head commit suicide fall asleep
ϕ-coefficient	bridge gap shrug shoulder press button		Distributed freq. of object (DFO) [27]	come true become difficult make sure
Nearest MI (NMI) [18]	Collocations with no neigh. MI		Whether object can occur as a verb	(Binary feature)
Whether object is a nomin. of some verb	(Binary feature)			

Table 2. List of features and their top-3 example collocations

6 Features

Each collocation is represented by a vector whose dimensions are the statistical features obtained from the British National Corpus. This list of features are given in Table 2.[2] While conducting the experiments, all features are scaled from 0 to 1 to ensure that all features are represented uniformly.

7 Experiments - Classification

The evaluation data (765 vectors) is divided randomly into training and testing vectors in 10 ways for cross-validation. The training data consists of 90% of 786 vectors and the testing data consists of the remaining.

We used various Machine Learning techinques to classify the V-N collocations into MWEs and non-MWEs. For every classifier, we calculated the average accuracy of all the test sets of each of the annotators. We then compare the average accuracies of all the classifiers. We found that the classifier that we used, the technique of weighted features in distance-weighted nearest-algorithm, performs somewhat better than other machine learning techniques.

[2] The formulas of features are not given due to lack of space.

The following are brief descriptions of the classifiers that we used in this paper.

7.1 Nearest-neighbour algorithm

This is an instance-based learning technique where the test vector is classified based on its nearest vectors in the training data. The simple distance between two vectors x_i and x_j is defined as d(x_i,x_j), where

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2}.$$

Here, x is an instance of a vector and $a_r(x)$ is the value of the r^{th} feature.

One can use K neighbours to judge the class of the test vector. The test vector is assigned the class of maximum number of neighbours. This can be furthur modified by calculating the inverse weighted distance between the test vector and the neighbouring training vectors in each of the classes. The test vector is then assigned the class which has the higher inverse-weighted distance. One can also use all the training vectors and the weighted-distance principle to classify the test vector.

The average classification accuracy of each of the above methods on the test sets of each of the annotators is shown in Table 3.

	Simple K-Nearest neighbour				Weighted-distance Nearest neighbour			
Type	K=1	K=2	K=3		K=1	K=2	K=3	K=All
Annot.1	62.35	61.31	62.48		62.35	62.35	62.61	66.66
Annot.2	57.64	54.10	60.89		57.64	57.64	60.37	63.52

Table 3. Average accuracies of MWE recognition using simple nearest-neighbour algorithms and weighted distance nearest neighbour algorithms

7.2 SVM-based classifiers

SVMs [15] have been very successful in attaining high accuracy for various machine-learning tasks. Unlike the error-driven algorithms (Perceptron etc.), SVM searches for the two distinct classes and maximizes the margin between two classes. Data of higher dimension can also be classified using the appropriate Kernel. We used Linear and Polynomial Kernel (degree=2) to test the evaluation data. We also used the radial-basis network in SVMs to compare the results because of their proximity to the nearest-neigbour algorithms.

The average classification accuracy of each of the above methods on the test sets of each of the annotators is shown in Table 4.

	Linear Ker.	Polynomial Ker.		Radial Basis networks			
Parameters				$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 1.5$	$\sigma = 2.0$
Annot.1	65.89	65.75		67.06	66.66	66.93	67.06
Annot.2	62.61	65.09		64.17	63.51	62.99	62.99

Table 4. Average accuracies of MWE recognition using SVMs (Linear, Polynomial and Radial Basis Function Kernel)

7.3 Weighted features in distance-weighted nearest-neighbour algorithm

Among all the features used, only a few might be very relevant to recognizing the non-compositionality of the MWE. As a result, the distance metric used by the nearest-neighbour algorithm which depends on all the features might be misleading. The distance between the neighbour will be dominated by large number of irrelevant features.

A way of overcoming this problem is to weight each feature differently when calculating the distance between the two instances. This also gives us an insight into which features are mainly responsible for recognizing the non-compositionality of MWEs. The j^{th} feature can be multiplied by the weight z_j, where the values of $z_1 ... z_n$ are chosen to minimize the true classification error of the learning algorithm [20]. The distance using these weights is represented as

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (z_r * (a_r(x_i) - a_r(x_j)))^2}$$

, where z_r is the weight of the r^{th} feature.

The values of $z_1 ... z_n$ can be determined by cross-validation of the training data. We use leave-one-out cross-validation [21], in which the set of m training vectors are repeatedly divided into a training set of m-1 and a test set of 1, in all possible ways. So, each vector in the training data is classified using the remaining vectors. The classification accuracy is defined as

$$Clacc = 100 * (\sum_{1}^{m} classify(i)/m)$$

where classify(i)=1, if the i^{th} training example is classified correctly using the distance-weighted nearest neighbour algorithm, otherwise classify(i)=0.

Now, we try to maximize the classification accuracy in the following way,

- In every iteration, vary the weights of the features one by one.
- Choose the feature and its weight which brings the maximum increase in the value of *Clacc*. One can also choose the feature and its weight such that it brings the minimum increase in the value of *Clacc*.
- Update the weight of this particular feature and go for the next iteration.

– If there is no increase in classification accuracy, stop.

When the weights are updated such that there is maximum increase in classification accuracy in every step, the average accuracies are **66.92%** and **64.30%** on the test sets of the two annotators respectively. But when the weights are updated such there is a minimum increase in classification accuracy at every step, the average accuracies are **66.13%** and **64.04%** on the test sets of the two annotators respectively, which are slightly better than that obtained by the other Machine Learning Techniques.

In the above two methods (Updating weights such that there is maximum or minimum increase in classification accuracy), we add the weights of the features of each of the evaluation sets. According to the average weights, the top three features (having high average weight) are shown in Tables 5 and 6.

Annotator1	Weight	Annotator2	Weight
DFO	1.09	MI	1.17
T-Score	1.0	T-Score	1.1
Z-Score	1.0	ϕ-coefficient	1.0

Table 5. The top three features according to the average weight when there is maximum increase in Clacc at every step

Annot.1	Weight	Annot.2	Weight
DFO	1.07	MI	2.06
NMI	1.02	T-Score	1.0
Log-Like.	0.97	ϕ-coefficient	1.0

Table 6. The top three features according to the average weight calculated when there is minimum increase in Clacc at every step

In both the above cases, we find that the properties 'Mutual-Information' and the compositionality oriented feature 'Distributed Frequency of an Object' performed significantly better than the other features.

8 Experiments - Ranking

All the statistical measures show that the expressions ranked higher according to their decreasing values are more likely to be non-compositional. We compare these ranks with the average of the ranks given by the annotator (obtained from his rating). To compare, we use Spearman Rank-Order Correlation Coefficient (r_s), defined as

$$r_s = \frac{(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum (R_i - \bar{R})^2 \sum (S_i - \bar{S})^2}}$$

where R_i is the rank of i^{th} x value, S_i is the rank of i^{th} y value, $\bar{R}$ is the mean of the R_i values and $\bar{S}$ is the mean of S_i values.

We use an SVM-based ranking system [16] for our training. Here, we use 10% of the 765 vectors for training and the remaining for testing. The SVM-based ranking system builds a preference matrix of the training vectors to learn. It then ranks the test vectors. The ranking system takes a lot of time to train itself, and hence, we decided to use only a small proportion of the evaluation set for training.

We also compare our ranks (the average of the ranks suggested by the classifier) with the gold standard using the Spearman Rank-Order Correlation Coefficient. The results are shown in Table 7.

MI	**-0.125**	Z-Score	-0.059
MI3	0.001	ϕ-coeff	-0.102
Log-Log	-0.086	DFO	**-0.113**
Log-Likelihood	0.005	NMI	**-0.167**
χ^2	-0.056	**Class.**	**0.388**
T-Score	0.045		

Table 7. The correlation values of the ranking of individual features and the ranking of classifier with the ranking of human judgements

In Table 7, we observe that the correlation between the ranks computed by the classifier and human ranking is better than the correlation between ranking of individual statistical features and human ranking.

We observe that among all the statistical features the ranks based on the properties 'Mutual Information', 'Distributed Frequency of an Object' [27] and 'Nearest mutual information' [18] correlated better with the ranks provided by the annotator. This is in accordance with the observation we made while describing the classification experiments, where we observed that the properties 'Distributed Frequency of an Object' and 'Mutual Information' contributed much to the classification of the expressions. When we compare the correlation values of MI, Log-likelihood and χ^2, we see that the Mutual-Information values correlated better. This result is similar to the observation made by McCarthy, Keller and Caroll [19] for phrasal verbs.

9 Conclusion

In this paper, we integrated the statistical features using various classifiers and investigated their suitability for recognising non-compositional MWEs of the V-N type. We also used a classifier to rank the V-N collocations according to their relative compositionality. This type of MWEs constitutes a very large percentage of all MWEs and are crucial for NLP applications, especially for Machine Translation. Our main results are as follows.

- The technique of weighted features in distance-weighted nearest neighbour algorithm performs better than other Machine Learning Techniques in the task of recognition of MWEs of V-N type.
- We show that the correlation between the ranks computed by the classifier and human ranking is significantly better than the correlation between ranking of individual features and human ranking.
- The properties 'Distributed frequency of object' and 'Nearest MI' contribute greatly to the recognition of the non-compositional MWEs of the V-N type and to the ranking of the V-N collocations based on their relative compositionality.

Our future work will consist of the following tasks

- Evaluate the effectiveness of the techniques developed in this paper for applications like Machine Translation.
- Improve our annotation guidelines and create more annotated data.
- Extend our approach to other types of MWEs.

Acknowledgements

We want to thank Libin Shen and Nikhil Dinesh for their help in clarifying various aspects of Machine Learning Techniques. We would like to thank Roderick Saxey and Pranesh Bhargava for annotating the data and Mark Mandel for considerable editorial help.

References

1. Abeille, Anne . Light verb constuctions and extraction out of NP in a tree adjoining grammar. Papers of the 24th Regional Meeting of the Chicago Linguistics Society. (1988)
2. Akimoto, Monoji . Papers of the 24th Regional Meeting of the Chicago Linguistics Society. Shinozaki Shorin . (1989)
3. Baldwin, Timothy and Bannard, Colin and Tanaka, Takaaki and Widdows, Dominic . An Empirical Model of Multiword Expression . Proceedings of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment. (2003)
4. Bannard, Colin and Baldwin, Timothy and Lascarides, Alex . A Statistical Approach to the Semantics of Verb-Particles . Proceedings of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment. (2003)
5. Bikel, Daniel M. . A Distributional Analysis of a Lexicalized Statistical Parsing Model . Proceedings of EMNLP . (2004)
6. Becker, Joseph D. . The Phrasal Lexicon . Theoritical Issues of NLP, Workshop in CL, Linguistics, Psychology and AI, Cambridge, MA. (1975)
7. Breidt, Elisabeth . Extraction of V-N-Collocations from Text Corpora: A Feasibility Study for German . CoRR-1996 . (1995)
8. Church, K. and Gale, W. and Hanks, P. and Hindle, D. . Parsing, word associations and typical predicate-argument relations . Current Issues in Parsing Technology. Kluwer Academic, Dordrecht, Netherlands, 1991 . (1991)

9. Church, K. and Patrick Hanks . Word Association Norms, Mutual Information, and Lexicography . Proceedings of the 27th. Annual Meeting of the Association for Computational Linguistics, 1990 . (1989)
10. Dunning, Ted . Accurate Methods for the Statistics of Surprise and Coincidence . Computational Linguistics - 1993 . (1993)
11. Stefan Evert and Brigitte Krenn . Methods for the Qualitative Evaluation of Lexical Association Measures . Proceedings of the ACL - 2001 . (2001)
12. Charles Fillmore . An extremist approach to multi-word expressions . A talk given at IRCS, University of Pennsylvania, 2003. (2003)
13. Fontenelle and Bruls, Th. W. and Thomas, L. and Vanallemeersch, T. and Jansen, J. . Survey of collocation extraction tools . Deliverable D-1a, MLAP-Project 93-19 DECIDE, University of Liege, Belgium. (1994)
14. Diaz-Galiano, M.C. and Martin-Valdivia, M.T. and Martinez-Santiago, F. and Urena-Lopez, L. A. . Multi-word Expressions Recognition with the LVQ Algorithm. Proceedings of Methodologies and Evaluation of Multiword Unit in Real-world Applications, LREC, 2004 . (2004)
15. Joachims, T. . Making large-Scale SVM Learning Practical . Advances in Kernel Methods - Support Vector Learning . (1999)
16. Joachims, T. . Optimizing Search Engines Using Clickthrough Data. Advances in Kernel Methods - Support Vector Learning edings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM, 2002. (2002)
17. Kilgariff, A. and Rosenzweig, J. . Framework and Results for English Senseval . Computers and the Humanities, 2000 . (2000)
18. Dekang Lin . Automatic Identification of non-compositonal phrases. Proceedings of ACL- 99, College Park, USA . (1999)
19. McCarthy, D. and Keller, B. and Carroll, J. . Detecting a Continuum of Compositionality in Phrasal Verbs . Proceedings of the ACL-2003 Workshop on Multi-word Expressions: Analysis, Acquisition and Treatment, 2003. (2003)
20. Mitchell, T. Instance-Based Learning . Machine Learning, McGraw-Hill Series in Computer Science, 1997 . (1997)
21. Moore, A. W. and Lee, M.S. . Proceedings of the 11 International Conference on Machine Learning, 1994. (1994)
22. Nunberg, G. and Sag, I. A. and Wasow, T. . Idioms . Language, 1994 . (1994)
23. Sag, I. A. and Baldwin, Timothy and Bond, Francis and Copestake, Ann and Flickinger, Dan. . Multi-word expressions: a pain in the neck for nlp . Proceedings of CICLing , 2002 . (2002)
24. Schone, Patrick and Jurafsky, Dan. Is Knowledge-Free Induction of Multiword Unit Dictionary Headwords a Solved Problem? . Proceedings of EMNLP , 2001 . (2001)
25. Schuler, William and Joshi, Aravind K. Relevance of tree rewriting systems for multi-word expressions. To be published. (2005)
26. Smadja, F. . Retrieving Collocations from Text : Xtract . Computational Linguistics - 1993 . (1993)
27. Tapanainen, Pasi and Piitulaine, Jussi and Jarvinen, Timo Idiomatic object usage and support verbs . 36th Annual Meeting of the Association for Computational Linguistics . (1998)
28. Venkatapathy, Sriram and Joshi, Aravind K. Recognition of Multi-word Expressions: A Study of Verb-Noun (V-N) Collocations. Proceedings of the International Conference on Natural Language Processing, 2004. (2004)